Homework 11

Due: Mon 30 Apr 2018, 3:00pm

Please follow these guidelines while writing up your CS 155 homeworks.

A major goal of this class is that you learn how to mathematically analyze probabilistic processes and events. Hence make sure that your proofs are rigorous, e.g., that every step is adequately justified. Try to keep your proofs as simple and as concise as possible, keeping in mind that you will potentially have to examine multiple proof strategies in order to achieve this. Make sure to CLEARLY STATE YOUR ASSUMPTIONS at the beginning of any solution, when necessary. As we are doing probabilistic analysis, we shall require that you clearly define the probability space over which you are working with and identify the events within it that you will analyse.

You are allowed to discuss with other students about these problems but you have to leave any meeting with other students without any written material about the problems, and the write ups should be written by yourself without help from others. If you have any questions regarding how these guidelines apply to a particular problem or what they mean in general please post on the newsgroup or email the TAs.

Problem 1

Read Section 14.5.2 in the book and answer the following questions in detail and in your own words (don’t just copy a few lines from the book!):

a. Define the $(\epsilon, \delta)$-approximation problem for frequent itemsets.

b. Prove a sample size bound for the $(\epsilon, \delta)$-approximation problem using the Chernoff and union bounds.

c. Improve the above bound on sample size when all the transactions have size bounded by $\ell$.

d. Formulate the $(\epsilon, \delta)$-approximation for frequent itemsets as an $\epsilon$-sample problem. What is the range space?

e. Answer exercise 14.12. Hint: for 14.12b, let $q_t$ be the largest integer $q_t$ such that out of the first $t$ transactions, there are at least $q_t$ transactions each of length at least $q_t$. Construct an algorithm that, after
having seen $t$ transactions, stores the maximum $q_t$ transactions with length longer than $q_t$. 